PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68 — A1

(11) International Publication Number: **WO 99/05320**

(43) International Publication Date: 4 February 1999 (04.02.99)

(21) International Application Number: PCT/US98/15041

(22) International Filing Date: 21 July 1998 (21.07.98)

(30) Priority Data:
60/053,436 22 July 1997 (22.07.97) US

(71) Applicant *(for all designated States except US)*: RAPIGENE, INC. [US/US]; 1631–220th Street S.E., Bothell, WA 98021 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: TABONE, John, C. [US/US]; 12117 Northeast 166th Place, Bothell, WA 98011 (US). VAN NESS, Jeffrey [US/US]; 10020 49th Avenue Northeast, Seattle, WA 98125 (US). MOYNIHAN, Kristen [US/US]; 5026 9th Avenue Northeast, Seattle, WA 98105 (US).

(74) Agents: PARKER, David, W. et al.; Seed and Berry LLP, 6300 Columbia Center, 701 Fifth Avenue, Seattle, WA 98104–7092 (US).

(81) Designated States: AL, AM, AT, AU, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: MULTIPLE FUNCTIONALITIES WITHIN AN ARRAY ELEMENT AND USES THEREOF

(57) Abstract

The present invention provides arrays of oligonucleotides on a solid substrate wherein a discrete area has at least two oligonucleotides with different sequences. These arrays are useful in hybrization assays, especially in conjunction with cleavable mass spectrometry tags.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

MULTIPLE FUNCTIONALITIES WITHIN AN ARRAY ELEMENT AND USES THEREOF

TECHNICAL FIELD

This invention relates generally to solid substrates with arrays of oligonucleotides printed on their surfaces, and in particular, to arrays with multiple oligonucleotides of differing sequences in a discrete area of the array.

BACKGROUND OF THE INVENTION

Replicate arrays of biological agents have been used to facilitate parallel testing of many samples. For example, sterile velvet cloths and a piston-ring apparatus has long been used to make replicates of bacterial and yeast colonies to agar plates each containing a different growth medium, as a means of rapidly screening a large number of independent colonies for different growth phenotypes (Lederberg and Lederberg, *J. Bacteriol. 63* :399, 1952). Likewise, 96-well microtiter plates are used to organize and store in an easily accessed fashion large numbers of *e.g.* cell lines, virus isolates representing recombinant DNA libraries, or monoclonal antibody cell lines.

The advent of large scale genomic projects and the increasing use of molecular diagnostics has necessitated the development of large volume throughput methods for screening nucleic acids. Recently, methods have been developed to synthesize large arrays of short oligodeoxynucleotides (ODNs) bound to a glass or silicon surface that represent all, or a subset of all, possible nucleotide sequences (Maskos and Southern, *Nucl. Acids Res. 20*: 1675, 1992). These ODN arrays have been made used to perform DNA sequence analysis by hybridization (Southern et al., *Genomics 13*: 1008, 1992; Drmanac et al., *Science 260*: 1649, 1993), determine expression profiles, screen for mutations and the like. For all these uses, the number of oligonucleotides needed is large, and thus high density arrays (>1000 oligonucleotides per 1 cm^2) have been developed. However, it would be advantageous in terms of time and economics to use lower density arrays. Currently, such arrays are limited by the

means of attaching the oligonucleotides (often in situ synthesis) and the variety of detectable markers.

The present invention discloses methods and compositions for producing arrays that have more than one nucleotide sequence per discrete area, and further provides other related advantages.

SUMMARY OF THE INVENTION

Within one aspect of the present invention, arrays of oligonucleotides are provided comprising a solid substrate with a surface comprising discrete areas of nucleic acid molecules, preferably oligonucleotides, wherein at least one area contains at least two nucleic acid molecules selected to have different sequences. Preferably, there are less than 1000 discrete areas. Also preferably, each area contains at least two oligonucleotides with different sequence and more preferably, at least one area contains from 2 to about 100 different oligonucleotide sequences.

In certain embodiments, an area of the array is from about 20 to about 500 microns in diameter, and wherein a center to center distance between areas is from about 50 to 1500 microns.

In preferred embodiments, the oligonucleotides have known sequences. In other preferred embodiments, the oligonucleotides are covalently attached to the surface of the substrate, preferably through an amine linkage, such as poly(ethyleneimine).

In another aspect, the invention provides a method of hybridization analysis, comprising (a) hybridizing labeled nucleic acid molecules to the array of oligonucleotides according to claim 1; and (b) detecting label in areas of the array, therefrom determining which oligonucleotides on the array hybridized. In preferred embodiments, at least one area of the array contains an oligonucleotide of known sequence and one of the labeled nucleic acid molecules is complementary to the oligonucleotide. Preferably, the labeled nucleic acid molecules comprise nucleic acid molecules with different sequences, each carrying a different label. Such labels may be

selected from the group of radioactive molecules, fluorescent molecules, and cleavable mass-spec tags.

In another aspect, the invention provides a method of identifying nucleic acid molecules in a sample, comprising (a) hybridizing labeled oligonucleotides to the nucleic acid molecules to form duplexes; (b) isolating the duplexes; (c) denaturing the duplexes; (d) hybridizing the labeled oligonucleotides to the array of oligonucleotides described herein, wherein the oligonucleotides on the array are complementary to the labeled oligonucleotides; and (e) detecting label in areas of the array; therefrom identifying the nucleic acid molecules in the sample.

In yet another aspect, methods are provided for identifying nucleic acid molecules in a sample, comprising (a) hybridizing oligonucleotides to the nucleic acid molecules; (b) extending the oligonucleotide in the presence of a single labeled nucleotide to form duplexes; (c) denaturing the duplexes; (d) hybridizing the labeled oligonucleotides to the array of oligonucleotides as described herein, wherein the oligonucleotides on the array are complementary to the labeled oligonucleotides; and (e) detecting label in areas of the array; therefrom identifying the nucleic acid molecules in the sample. In preferred embodiments, the oligonucleotides in a discrete area of the array are complementary to extension products of the nucleic acid molecules.

In yet another aspect, the invention provides a method of identifying nucleic acid molecules in a sample, comprising (a) hybridizing at least two oligonucleotides to the nucleic acid molecules to form duplexes, wherein at least one oligonucleotide is labeled and the oligonucleotides hybridize to adjacent sequences on the nucleic acid molecules; (b) ligating the oligonucleotides; (c) denaturing the duplexes; (d) hybridizing the labeled oligonucleotides to the array of oligonucleotides described herein, wherein the oligonucleotides on the array are complementary to the ligated oligonucleotides; and wherein hybridization does not occur to unligated oligonucleotides; and (e) detecting label in areas of the array; therefrom identifying the nucleic acid molecules in the sample.

In yet another aspect, methods are provided for identifying mRNA molecules in a sample, comprising (a) hybridizing labeled oligonucleotides to the

mRNA molecules to form duplexes; (b) isolating the duplexes; (c) denaturing the duplexes; (d) hybridizing the labeled oligonucleotides to the array of oligonucleotides described herein, wherein the oligonucleotides on the array are complementary to the labeled oligonucleotides; and (e) detecting label in areas of the array; therefrom identifying the mRNA molecules in the sample. In preferred embodiments, the mRNA molecules are isolated from cells treated with compounds suspected of being toxins. In other preferred embodiments, the oligonucleotides on the array are sequences of cytokines.

These and other aspects of the present invention will become evident upon reference to the following detailed description and attached drawings. In addition, various references are set forth below which describe in more detail certain procedures or compositions (*e.g.*, plasmids, etc.), and are therefore incorporated by reference in their entirety.

The biomolecule arrays of the present invention may contain, or be used in conjunction with, tagged biomolecules, for example, oligonucleotides covalently bonded to cleavable tags. These tagged biomolecules may be used in methods of the present invention, and assay procedures such as oligonucleotide sequencing and gene expression assays, among others. Exemplary tagged biomolecules, and assays which may use the same, are described in U.S. Patent Application Nos. 08/786,835; 08/786,834 and 08/787,521, each filed on January 22, 1997, as well as in three U.S. continuation-in-part patent applications having Application Nos. 08/898,180; 08/898,564; and 08/898,501, each filed July 22, 1997; and in PCT International Publication Nos. WO 97/27331; WO 97/27325; and WO 97/27327. These six U.S. Patent Applications and three PCT International Publications are each hereby fully incorporated herein by reference in their entireties.

The biomolecule arrays of the present invention may also be used in performing amplification and other enzymatic reactions, as described in U.S. Provisional Patent Application No. 60/053,428 titled "Amplification And Other Enzymatic Reactions Performed On Nucleic Arrays" as filed July 22, 1997, and like-

titled U.S. Non-Provisional Patent Application No. _____ filed concurrently herewith, both being fully incorporated herein by reference in their entireties.

The biomolecule arrays of the present invention, and arrays useful in the methods of the present invention, may be prepared according to techniques disclosed in, for example, U.S. Provisional Patent Application No. 60/053,435 titled "Apparatus And Methods For Arraying Solution Onto A Solid Support" as filed July 22, 1997, and like-titled U.S. Non-Provisional Patent Application No. _____ filed concurrently herewith, both being fully incorporated herein by reference in their entireties.

The biomolecule arrays of the present invention, and arrays useful in the methods of the present invention, may be prepared according to techniques disclosed in, for example, U.S. Provisional Patent Application No. 60/053,352 titled "Polyethylenimine-Based Biomolecule Arrays" as filed July 22, 1997, and like-titled U.S. Non-Provisional Patent Application No. _____ filed concurrently herewith, both being fully incorporated herein by reference in their entireties.

Computer systems and methods for correlating data, as disclosed in, for example, U.S. Provisional Patent Application No. 60/053,429 titled "Computer Method and System for Correlating Data" as filed July 22, 1997, and like-titled U.S. Non-Provisional Patent Application No. _____ filed concurrently herewith (both being fully incorporated herein by reference in their entireties) may be used in conjunction with the biomolecule arrays and methods as described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows photomicrographs of arrayed microspheres taken under visible light illumination (top panel) and fluorescence illumination (bottom panel).

Figure 2 shows a CCD camera image of an array produced by a robot using the methodology of the invention, where the domains are approximately 100-150 microns in average diameter with 200 micron center to center spacing between spots. The standard deviation of spot diameter is approximately 15%.

Figure 3 shows an array of microspots prepared according to the invention and developed using Vector Blue (Vector Laboratories, Burlingame, California) and imaged with a CCD camera and microscope.

Figure 4 is an illustration showing how two different oligonucleotides, both present within a single array element, may be identified and partially quantified according to the present invention.

DETAILED DESCRIPTION OF THE INVENTION

As noted above, the present invention provides arrays with multiple sequences in a single discrete area. In certain embodiments, the invention provides more than 1, and preferably 10 to 100 different oligonucleotide sequences or polynucleotide sequences in a single element within an array. In the case of oligonucleotides, between 2 and approximately 100 oligonucleotides would be synthesized individually on a commercial synthesizer, combined and printed as a single element in a discrete area of an array. Nucleic acids that are double stranded, single stranded, comprising DNA, RNA, or both may be coupled to a solid substrate. Double-stranded molecules may be generated by amplification, enzymatic digestion, or the like. Essentially, any nucleic acid molecule that has a primary amine group (for coupling to polyethylenimine) or other reactive group can be conjugated to the arrays within the present invention. Also within the present invention, the sequence of every individual nucleic acid within a particular area may be unknown.

As used herein, an array refers to a collection of oligonucleotide or polynucleotide sequences that are placed on a solid support in discrete areas. Preferably, the areas form some identifiable pattern or regular intervals. An array is typically composed of 2 to 1000 elements, but may be composed of over 1000 elements (discrete areas) as well. Each area is separated by some distance in which no nucleic acid or oligonucleotide is bound or deposited. Typical area sizes are 20 to 500 microns and typical center to center distances of the area range from 50 to 1500 microns.

This invention also describes the use of multiple sequences within a single element within an array. The method permits the use of low element arrays (i.e.,

10 elements to for example 400 elements) for many purposes. For example, these combined sequence, low element arrays can be used for pathogen identification, profiling determinations, toxicology testing and the like.

I. APPLICATION OF TEMPLATES TO SOLID SUBSTRATE

A. Substrate preparation

A substrate for arrays is prepared from a suitable material. The substrate is preferably rigid and preferably has a surface that is substantially flat. In some embodiments, the surface may have raised portions to delineate regions. Typical substrates are silicon wafers and borosilicate slides (*e.g.*, microscope glass slides), although other materials known in the art may be substituted. An example of a particularly useful solid support is a silicon wafer that is typically used in the electronics industry in the construction of semiconductors. The wafers are highly polished and reflective on one side and can be easily coated with various linkers, such as poly(ethyleneimine) using silane chemistry. Wafers are commercially available from companies such as WaferNet, San Jose, CA.

Nucleic acid molecules or other biopolymers, such as peptides, may be synthesized, generated or isolated and applied to the substrate. Nucleic acids and peptides may be synthesized in an automated fashion using a commercially available machine. In preferred embodiments, the molecules are deposited on the solid substrate and are covalently attached to the substrate.

In certain embodiments, the surface of the substrate is prepared for the oligonucleotides. The surface may be prepared by, for example, coating with a chemical that increases or decreases the hydrophobicity or coating with a chemical that allows covalent linkage of the nucleic acid molecules or other polymeric sequences. Some chemical coatings may both alter the hydrophobicity and allow covalent linkage. Hydrophobicity on a solid substrate may readily be increased by silane treatment or other treatments known in the art. A chemical that allows covalent linkage is generally referred to as a linker. These linker molecules adhere to the surface of the substrate and comprise a functional group that reacts with biomolecules. Many such linkers are

readily available. For example, solid supports are modified with photolabile-protected hydroxyl groups (*see,* U.S. Patent Nos.5,412,087; 5,571,639; 5,593,839), alkoxy or aliphatic derivatized hydroxyl groups (U.S. Patent No. 5,436,327), or other chemicals (*see e.g.*, U.S. Patent No. 5,445,934; EP Patent No. EP-B1-0,373,203; U.S. Patent No. 5,474,796; U.S. Patent No. 5,202,231).

A preferred coating that both decreases hydrophobicity and provides linkers is poly(ethyleneimine). In addition, poly(ethyleneimine) (PEI) coated solid substrates have the benefit of long shelf life stability. The coating of silicon wafers and glass slides with polymers such as poly(ethyleneimine) can be performed in-house or through companies such as Cel Associates (Houston, Texas). Glass slides can also be coated with a reflective material or coated with PEI using silane chemistry. The PEI coating permits the covalent attachment of single or double stranded oligonucleotides, single or double stranded long DNA molecules or fragments or any other amine-containing biomolecules to the solid support. Oligonucleotides may be covalently attached at the 5' using a hexylamine modification, which places a primary amine at the 5'-end of the oligonucleotide. The 5'-amine on the oligonucleotide may then be reacted with a cross-linker, such that the oligonucleotide is covalently attached to the polymer coating on the solid support.

Any nucleic acid type can be covalently attached to a PEI coated surface as long as the nucleic acid contains a primary amine. Amplified products (*e.g.*, by PCR) may be modified to contain a primary amine by using 5'-hexylamine-conjugated primers. Amine groups may be introduced into amplified products and other nucleic acid duplexes by nick translation using allyl-dUTP (Sigma, St. Louis, MO). As well, amines may be introduced into nucleic acids by polymerases, such as terminal transferase, or by ligation of short amine-containing oligonucleotides. Other suitable methods known in the art may be substituted.

Cross linkers suitable for amine groups are generally commercially available (*see, e.g.*, Pierce, Rockford, IL). A typical cross-linker is trichlorotriazine (cyanuric chloride) (Van Ness et al., *Nucleic Acids Res. 19:* 3345-3350, 1991). Briefly, an excess of cyanuric chloride is added to the oligonucleotide solution (*e.g.*, a 10 to

1000-fold molar excess of cyanuric chloride over amines) at a typical oligonucleotide concentration of 0.01 to 1 μg/ml, and preferably about 0.1 μg/ml. The reaction is buffered using common buffers such as sodium phosphate, sodium borate, sodium carbonate, or Tris HCL at a pH range from 7.0 to 9.0. The preferred buffer is freshly prepared 0.2 M NaBorate at pH 8.3 to pH 8.5. Ten μl of 15 mg/ml solution of cyanuric chloride is added and allowed to react with constant agitation from 1 to 12 hours and preferably approximately 1 hour. Reaction temperature may range from 20 to 50°C with the preferred reaction temperature at 25°C (or ambient temperature).

When cyanuric chloride is used as a cross linker, there is no need to remove the excess crosslinker prior to printing the nucleic acids on a solid substrate. Excess cyanuric chloride in the reaction mixture does not interfere or compete with the covalent attachment of the nucleic acid or oligonucleotides to the PEI coated solid support, because of an excess of amines on the solid support over the number of cyanuric chloride molecules. In a preferred embodiment, cross-linked oligonucleotides are not purified prior to the printing step.

If the nucleic acids or other amine-containing polymers are to be covalently attached, the activated polymers are allowed to react with the solid support for 1 to 20 hours at 20 to 50°C and preferably for 1 hour at 25°C. The free amines on the solid support are then capped to prevent non-specific attachment of other nucleic acids. Capping is accomplished by reacting the solid support with 0.1 to 2.0 M succinic anhydride, and preferably 1.0 M succinic anhydride in 70% m-pyrol and 0.1 M NaBorate, for 15 minutes to 4 hours with a preferred reaction time of 30 minutes at 25°C. The solid support is then incubated in a 0.1 to 10.0 M NaBorate, pH 7 to pH 9 (preferably 0.1 M NaBorate pH 8.3) solution containing 0.1 to 5 M glycine (preferably 0.2 M glycine) and then washed with detergent-containing solution. This "caps" any dichloro-triazine that may be covalently bound to the PEI surface. Preferably, the solid support is further heated to 95°C in 0.01 M NaCl, 0.05 M EDTA and 10 mM Tris pH 8.0 for 5 minutes to remove any non-covalently attached nucleic acids. In the case where double stranded nucleic acids are printed onto a solid substrate, this step also converts (denatures) the double strand to a single strand form.

In the currently used array formats, the arrays contain the lowest possible information content: each element in the array corresponds to just one sequence. Therefore, every element in the array is of known sequence and if an element scores positively in an array (*e.g.*, hybridizes), the sequence of the hit is known. In essence, there is a one to one correspondence of the material contained in the element of the array and the information content contained within the array.

In the arrays of the present invention, additional information content is provided by having multiple sequences in each element. In the simplest form, each area of the array contains a unique sequence plus a control sequence for measuring the amount of material per element in the array. If two oligonucleotides are immobilized within a single element, one of the oligonucleotides can be used for the means of conducting quality control or quality assurance on the array. The "control" oligonucleotide sequence may serve as a capture site for the complementary oligonucleotide which contains or possesses some label that is detectable. The "control" oligonucleotide may also serve as an internal control for the arraying process and method.

In this format, as well as the formats described below, at least one area contains multiple nucleic acid sequences. In a preferred embodiment, each area contains multiple sequences. For some purposes, an intermediate value, that is, a fraction of areas have multiple sequences and the remaining fraction have a single sequence. As well, when multiple sequences are present, there are at least two sequences, generally not more than 1000, and preferably 2 to 100.

In preferred aspects, the arrays contain an intermediate amount of information content. For example, in one embodiment, the array contains two sequences per element. Thus, there is a 2:1 correspondence between information content per element and numbers of elements per array, but a loss of exact sequence identity using a single label in this format. Every element in the array is of composed of two possible known sequences and if an element scores positively in an array, the sequence of the hit is the possibility of two distinct sequences. The identity of the positive can be determined however by the use of multiple detection molecules. As

described below, the use of different fluorescent molecules, colored microbeads, radioactive molecules, chromogenic substrates, combinations of these markers, a combination of one of these markers and chemiluminescene, or cleavable mass-spectrometry tags can provide information of which sequences in the area of an array have been hybridized.

As shown in the table below, as the number of different sequences increases in each area of the array, either the number of repetitious reactions that need to be performed for unambiguous identification (deconvolution) increases or the number of different tags needed increases.

# sequences/ element	sequence known?	correspondence	deconvolution	# tags
1	yes	1:1	none	1
2	no	2:1	1/2	2
3	no	3:1	1/3	3
5	no	5:1	1/5	4
10	no	10:1	1/10	10
n	no	n:1	1/n	n

where n preferably ranges from 1-100 per element in an array. The number of sequences per element is the number of different sequences that placed in a single element per array. The sequence known column represents the ability to determine the sequence of the target (test) sample with a single label. The correspondence is the amount of information or the possibility that any given sequence is represented within a single element with an array or subarray. The deconvolution column represents the number of sub-assays that need to be performed in order to determine the exact sequence of the targets (test sample) which scored positively in the assay. For example, if a hit occurs in an area containing 10 different sequences, an unambiguous determination can be made by printing the 10 different sequences separately and hybridizing.

Substantial and myriad advantages are attained from using more than one sequence of nucleic acid or more than one oligonucleotide per element within an array. For example, families of related nucleic acid sequences can be place within a single

element of an array, and hence the gene activity of any test sample can be determined by examining the pattern of hybridization across the array. In the case of toxicology testing, an array could be built which would possess the following elements: an element containing sequences for proinflammatory cytokine induction (IL-1, IL-2, IL-6, IL-8), an element containing sequences for anti-inflammatory cytokine induction (IL-4, IL-12), an element containing sequences for lipid modifying enzymes (platelet-activating factor acetyl-transferase, (platelet-activating factor acetyl-hydrolase), and an element containing sequences for TNF (TNF-alpha and TNF-beta), and the like. One skilled in the art will recognize that the choice of nucleic acid sequences will depend, in part, upon what is being tested.

B. Methods of applying nucleic acid molecules to solid substrates

Oligonucleotides, nucleic acid molecules or other biopolymers are "printed" (delivered or applied) on a solid substrate. In preferred embodiments, the polymers are applied in a regular pattern or array.

A variety of printing methods are available for applying nucleic acids, such as oligonucleotides or DNA fragments, to a solid substrate in an array pattern. As a general guideline, the delivery mechanism must be capable of positioning very small amounts of liquids (*e.g.*, nanoliters) in small regions (*e.g.*, 300 μm diameter dots) where the regions are very close to one another (*e.g.*, 1 mm or less separation). Preferably the printing technique is amenable to automation. One such technique is ink-jet printing using multiple heads. Very fine pipettes may also be used. A preferred means of printing is using spring probes as described herein.

Sample pick-up, transfer and micro-droplet deposition is greatly enhanced when using a liquid transfer device that has a hydrophilic surface, especially when that device is a modified spring probe. Spring probes are made hydrophilic through the use of chemical agents acting to modify the surface of the probe or through coating the probe with a hydrophilic substance. In a preferred method, the tip of the spring probe is soaked in a 25 - 200 mM solution of 1,4-dithiothreitol, 0.1 M sodium

borate for 15 min to 2 hrs. Dithiothreitol reacts with gold surfaces through a thiol-gold coordination, which essentially hydroxylates the surface, making it hydrophilic.

The hydrophilic surface promotes an even coating of sample when the spring probe is dipped in solution. The fluted probe becomes evenly and consistently loaded with liquid drawn to the probe surface by its hydrophilic nature. Solutions with viscosity enhancing chemicals, such as glycerol, provide especially improved handling capabilities using hydrophilic surfaces. With these solutions, the glycerol adheres to the probe even as it pulled from the source of liquid. As a sample is transferred from its source to a solid support, the hydrophilic surface of the probe continues to benefit liquid handling by retaining the sample being transferred and inhibiting the sample from randomly dripping or running during transport. When a sample bearing spring probe comes into contact with a solid support, sample is deposited from the tip of the spring probe onto the surface of the solid support, especially in the case of a sample containing a viscosity enhancing solution. The size of the areas spotted generally range from 10-200 μm with a typical center to center distance of 25-500 μm.

Briefly, in a typical procedure, a solution of the nucleic acid is uniformly mixed in 57% glycerol and then printed onto the solid support. Within the context of this invention, the biopolymers may be either nucleic acid molecules or protein molecules. When nucleic acids are used, they may comprise single or double stranded DNA, single or double stranded RNA, oligonucleotides, hybrid DNA-RNA molecules or duplexes, PNA nucleic acids with a protein backbone and the like.

II. REACTION COMPONENTS AND CONDITIONS

As noted above, the present invention provides methods for hybridizations to the nucleic acid molecules on the solid substrate. As noted above, the nucleic acids may be covalently attached to the surface of the substrate or may be deposited on the substrate without attachment. Typically, the oligonucleotides are printed first and other reagents are subsequently added.

A. Reagents, buffers, cofactors, etc.

Each area of the array that undergoes hybridization has in addition to template nucleic acids, the appropriate labeled nucleic acids, buffers, cofactors, and the like. Hybridization conditions are well known (*see,* Ausubel, et al., *Current Protocols in Molecular Biology*, Greene Publishing, 1995; Sambrook et al., *Molecular Cloning: A Laboratory Approach*, Cold Spring Harbor Press, 1987) and well described for hybridizations using short pieces of nucleic acids.

In a preferred embodiment, a hybotrope may be added to improve annealing of an oligonucleotide primer to template (see U. S. Application Nos. 60/026,621 (filed September 24, 1996); 08/719,132 (filed September 24, 1996); 08/933,924 (filed September 23, 1997); 09/002,051 (filed December 31, 1997); and PCT International Publication No. WO 98/13527 which are all incorporated herein in their entireties). A hybotrope refers to any chemical that can increase the enthalpy of a nucleic acid duplex by 20% or more when referenced to a standard salt solution (*i.e.*, 0.165 M NaCl). A chemical exhibits hybotropic properties when, as a solution an 18 bp oligonucleotide duplex that is 50% G+C has a helical-coil transition (HCT) of 15°C or less. HCT is the difference between the temperatures at which 80% and 20% of the duplex is single-stranded. The temperature for annealing is then chosen to be the discrimination temperature, which is a temperature at which a hybridization reaction is performed that allows detectable discrimination between a mismatched duplex and a perfectly matched duplex. A range of temperatures satisfy criteria of a discrimination temperature.

III. DETECTION OF REACTION PRODUCTS

Reaction products may be detected by a variety of methods. Preferably, one of the reaction components is labeled. In amplification reactions, the oligonucleotide primers or the nucleotides are conveniently labeled. Preferably, the primers contain a label. In single nucleotide extension assay, the added nucleotide is generally labeled, in oligonucleotide ligation assay, one or more of the oligonucleotides

are labeled, in other synthesis reactions, either the primer or the nucleotides are typically labeled.

Commonly employed labels include, but are not limited to, biotin, fluorescent molecules, radioactive molecules, chromogenic substrates, chemiluminescence, and the like. The methods for biotinylating nucleic acids are well known in the art, as are methods for introducing fluorescent molecules and radioactive molecules into oligonucleotides and nucleotides.

When biotin is employed, it is detected by avidin, streptavidin or the like, which is conjugated to a detectable marker, such as an enzyme (*e.g.*, horseradish peroxidase) or radioactive label (*e.g.*, ^{32}P, ^{35}S, ^{33}P). Enzyme conjugates are commercially available from, for example, Vector Laboratories (Burlingame, CA). Steptavidin binds with high affinity to biotin, unbound streptavidin is washed away, and the presence of horseradish peroxidase enzyme is then detected using a precipitating substrate in the presence of peroxide and appropriate buffers. The product may be detected using a microscope equipped with a visible light source and a CCD camera (Princeton Instruments, Princeton, NJ). With such an instrument, an image of approximately 10,000 μM x 10,000 μM can be scanned at one time.

Detection methods are well known for the above described labels, such as fluorescent or radioactive labels. Fluorescent labels can be identified and quantitated most directly by their absorption and fluorescence emission wavelengths and intensity. A microscope/camera setup using a fluorescent light source is a convenient means for detecting fluorescent label. Radioactive labels may be visualized by standard autoradiography, phosphor image analysis or CCD detector. Other detection systems are available and known in the art. For labels such as biotin, radioactive, or fluorescent, the number of different reactions that can be detected at a single time is limited. For example, the use of four fluorescent molecules, such as commonly employed in DNA sequence analysis, limits analysis to four samples at a time. Essentially, because of this limitation, each reaction must be individually assessed when using these detector methods.

A more advantageous method of detection allows pooling of the sample reactions on at least one array and simultaneous detection of the products. By using a tag having a different molecular weight or other physical attribute in each reaction, the entire set of reaction products can be harvested together and analyzed. (see U.S. Patent Application Nos. 08/786,835; 08/786,834; and 08/787,521, each filed on January 22, 1997, U.S. Patent Application Nos. 08/898,180; 08/898,564; and 08/898,501, each filed July 22, 1997; and PCT International Publication Nos. WO 97/27331; WO 97/27325; and WO 97/27327). Briefly, a "tag" molecule is used as a label. As used herein, a "tag" refers to a chemical moiety which is used to uniquely identify a "molecule of interest", and more specifically refers to the tag variable component as well as whatever may be bonded most closely to it in any of the tag reactant, tag component and tag moiety.

A tag useful in the present invention possesses several attributes:

(1) It is capable of being distinguished from all other tags. This discrimination from other chemical moieties can be based on the chromatographic behavior of the tag (particularly after the cleavage reaction), its spectroscopic or potentiometric properties, or some combination thereof. Spectroscopic methods by which tags are usefully distinguished include mass spectroscopy (MS), infrared (IR), ultraviolet (UV), and fluorescence, where MS, IR and UV are preferred, and MS most preferred spectroscopic methods. Potentiometric amperometry is a preferred potentiometric method.

(2) The tag is capable of being detected when present at 10^{-22} to 10^{-6} mole.

(3) The tag possesses a chemical handle through which it can be attached to the MOI which the tag is intended to uniquely identify. The attachment may be made directly to the MOI, or indirectly through a "linker" group.

(4) The tag is chemically stable toward all manipulations to which it is subjected, including attachment and cleavage from the MOI, and any manipulations of the MOI while the tag is attached to it.

(5) The tag does not significantly interfere with the manipulations performed on the MOI while the tag is attached to it. For instance, if the tag is attached

to an oligonucleotide, the tag must not significantly interfere with any hybridization or enzymatic reactions (*e.g.*, amplification reactions) performed on the oligonucleotide.

A tag moiety that is intended to be detected by a certain spectroscopic or potentiometric method should possess properties which enhance the sensitivity and specificity of detection by that method. Typically, the tag moiety will have those properties because they have been designed into the tag variable component, which will typically constitute the major portion of the tag moiety. In the following discussion, the use of the word "tag" typically refers to the tag moiety (*i.e.*, the cleavage product that contains the tag variable component), however can also be considered to refer to the tag variable component itself because that is the portion of the tag moiety which is typically responsible for providing the uniquely detectable properties. In compounds of the formula T-L-X, the "T" portion contains the tag variable component. Where the tag variable component has been designed to be characterized by, *e.g.*, mass spectrometry, the "T" portion of T-L-X may be referred to as T^{ms}. Likewise, the cleavage product from T-L-X that contains T may be referred to as the T^{ms}-containing moiety. The following spectroscopic and potentiometric methods may be used to characterize T^{ms}-containing moieties.

Thus, within one aspect of the present invention, methods are provided for determining the identity of a nucleic acid molecule or fragment (or for detecting the presence of a selected nucleic acid molecule or fragment), comprising the steps of (a) hybridizing tagged nucleic acid molecules from one or more selected target nucleic acid molecules, wherein a tag is correlative with a particular nucleic acid molecule and detectable by non-fluorescent spectrometry or potentiometry, (b) washing away non-hybridized tagged nucleic acids; (c) cleaving the tags from the tagged molecules, and (d) detecting the tags by non-fluorescent spectrometry or potentiometry, and therefrom determining the identity of the nucleic acid molecules. Examples of such technologies include for example mass spectrometry, infra-red spectrometry, potentiostatic amperometry or UV spectrometry.

IV. USES

As noted above, the methods described herein are applicable for a variety of purposes. For example, the arrays of oligonucleotides may be used to control for quality of making arrays, for quantitation or qualitative analysis of nucleic acid molecules, for detecting mutations, for determining expression profiles, for toxicology testing, and the like.

A. Internal controls in hybridization or making of arrays

In this embodiment, each area of the array contains the same nucleic acid sequence in addition to any other sequences. Thus, one common sequence is located in each element. The common sequence may be used as a control for monitoring quality control or quality assurance for making of the array. The "control" sequence may serve as a capture site for the complementary oligonucleotide that contains or possesses a detectable label. As such, reproducibility of the amount of nucleic acids in each area can be determined. If necessary, results can be normalized to the control sequence. Control sequences can be incorporated into any of the uses described herein.

B. Probe quantitation or typing

In this embodiment, 2 to 100 oligonucleotides are immobilized per element in an array where each oligonucleotide in the element is a different or related sequence. Preferably, each element possesses a known or related set of sequences. The hybridization of a labeled probe to such an array permits the characterization of a probe and the identification and quantification of the sequences contained in a probe population.

A generalized assay format that may be used in the particular applications discussed below is a sandwich assay format. In this format, a plurality of oligonucleotides (*e.g.* 2 to 100) of known sequence are immobilized in the same element of the array. Each element thus possesses a known or related set of sequences. The immobilized oligonucleotide is used to capture a nucleic acid (e.g., RNA, rRNA, a PCR product, fragmented DNA) and then a signal probe is hybridized to a different portion of the captured target nucleic acid.

Another generalized assay format is a secondary detection system. In this format, the arrays are used to identify and quantify labeled nucleic acids that have been used in a primary binding assay. For example, if an assay results in a labeled nucleic acid, the identity of that nucleic acid can be determined by hybridization to an array. These assay formats are particularly useful when combined with cleavable mass spectrometry tags.

C. Mutation detection

The detection of diseases is increasingly important in prevention and treatments. While multifactorial diseases are difficult to devise genetic tests for, more than 200 known human disorders are caused by a defect in a single gene, often a change of a single amino acid residue that results in the disease (Olsen, *Biotechnology: An industry comes of age,* National Academic Press, 1986).

Analyses may be performed before the implantation of a fertilized egg (Holding and Monk, *Lancet 3*:532, 1989) or in cells exfoliated from the respiratory tract or the bladder in connection with health checkups (Sidransky et al., *Science 252*:706, 1991). Also, when an unknown gene causes a genetic disease, methods to monitor DNA sequence variants are useful to study the inheritance of disease through genetic linkage analysis.

Mutations involving a single nucleotide can be identified in a sample by physical, chemical, or enzymatic means. Generally, methods for mutation detection may be divided into scanning techniques, which are suitable to identify previously unknown mutations, and techniques designed to detect, distinguish, or quantify known sequence variants. Several scanning techniques for mutation detection have been developed based on the observation that heteroduplexes of mismatched complementary DNA strands, derived from wild type and mutant sequences, exhibit an abnormal migratory behavior.

The methods described herein may be used for mutation screening. One strategy for detecting a mutation in a DNA strand is by hybridization of the test sequence to target sequences that are wild-type or mutant sequences. A mismatched sequence has a destabilizing effect on the hybridization of short oligonucleotide probes

to a target sequence (see Wetmur, *Crit. Rev. Biochem. Mol. Biol., 26*:227, 1991). The test nucleic acid source can be genomic DNA, RNA, cDNA, or amplification of any of these nucleic acids. Preferably, amplification of test sequences is first performed, followed by hybridization with short oligonucleotide probes immobilized on an array. An amplified product can be scanned for many possible sequence variants by determining its hybridization pattern to an array of immobilized oligonucleotide probes.

A label, such as described herein, is generally incorporated into the final amplification product by using a labeled nucleotide or by using a labeled primer. The amplification product is denatured and hybridized to the array. Unbound product is washed off and label bound to the array is detected by one of the methods herein. For example, when cleavable mass spectrometry tags are used,

D. Expression profiles / differential display

Mammals, such as human beings, have about 100,000 different genes in their genome, of which only a small fraction, perhaps 15%, are expressed in any individual cell. The process of normal cellular growth and differentiation, as well as the pathological changes that arise in diseases like cancer, are all driven by changes in gene expression. Differential display techniques permit the identification of genes specific for individual cell types.

Briefly, in differential display, the 3' terminal portions of mRNAs are amplified and identified on the basis of size. Using a primer designed to bind to the 5' boundary of a poly(A) tail for reverse transcription, followed by amplification of the cDNA using upstream arbitrary sequence primers, mRNA sub-populations are obtained. The differential display method has the potential to visualize all the expressed genes (about 10,000 to 15,000 mRNA species) in a mammalian cell by using multiple primer combinations.

The hybridization of amplified cDNA to a plurality of oligonucleotides immobilized in the same element of the array permits the identification and quantification of the sequences amplified, and thus, the starting mRNA population. On the array, 2 to 100 oligonucleotides may be immobilized per element wherein each oligonucleotide in the element is a different or related sequence. The identification of

hybridizing sequences permit the expression profiling of sequences of the target nucleic acid from which the probe is produced. Expression profiling can be used to measure the regulation of genes and messenger RNAs (mRNAs) in response to cellular signals, stimuli, and the like. For example, arrays for toxicology testing may be constructed such that individual areas contain sequences complementary to various cytokines (see above) For toxicology testing, the stimuli are generally small organic compounds or molecules which are suspected to be toxins or where the toxicology of a small organic molecules is to be determined.

As disclosed herein, a high throughput method for measuring the expression of numerous genes (*e.g.*, 1-2000) is provided. Within one embodiment of the invention, methods are provided for analyzing the pattern of gene expression from a selected biological sample, comprising the steps of (a) amplifying cDNA from a biological sample using one or more tagged primers, wherein the tag is correlative with a particular nucleic acid probe and detectable by non-fluorescent spectrometry or potentiometry, (b) hybridizing amplified fragments to an array of oligonucleotides as described herein, (c) washing away non-hybridized material, and (d) detecting the tag by non-fluorescent spectrometry or potentiometry, and therefrom determining the pattern of gene expression of the biological sample.

Tag-based differential display, especially using cleavable mass spectrometry tags, on solid substrates allows characterization of differentially expressed genes. It is based on the principle that most mRNAs expressed in two or more cell types or samples of interest can be directly compared after amplification of partial cDNA sequences from subsets of mRNA. Briefly, three one-base anchored oligo-dT primers are used in combination with a series of arbitrary 13 base oligonucleotides to reverse transcribe and amplify the mRNAs from a cell or sample of interest. For monitoring the expression of 15,000 genes, at least nine different arbitrary primers are preferably used. For a complete differential display analysis of two cell populations or two samples of interest, at least 400 amplification reactions are performed. With tag-based differential display analysis of two cell types, at least 1500 amplification reactions are easily and quickly performed.

E. Single nucleotide extension assay

The primer extension technique may be used for the detection of single nucleotide in a nucleic acid template (Sokolov, *Nucleic Acids Res.*, *18:*3671, 1989). In its original format, 20 and 30 base oligonucleotides complementary to the known sequence of the cystic fibrosis gene were extended in the presence of a single labeled nucleotide and correctly identified a single nucleotide change within the gene. The technique is generally applicable to detection of any single base mutation (Kuppuswamy et al., *Proc. Natl, Acad. Sci. USA*, 88:1143-1147, 1991).

Briefly, this method first hybridizes a primer to a sequence adjacent to a known single nucleotide polymorphism. The primed DNA is then subjected to conditions in which a DNA polymerase adds a labeled dNTP, typically a ddNTP, if the next base in the template is complementary to the labeled nucleotide in the reaction mixture. In a modification, cDNA is first amplified for a sequence of interest containing a single-base difference between two alleles. Each amplified product is then analyzed for the presence, absence, or relative amounts of each allele by annealing a primer that is 1 base 5' to the polymorphism and extending by one labeled base (generally a dideoxynucleotide). Only when the correct base is available in the reaction will a base to incorporated at the 3'-end of the primer. Extension products are then analyzed by hybridization to an array of oligonucleotides such that a non-extended product will not hybridize.

Briefly, in the present invention, each dideoxynucleotide is labeled with a unique tag. Of the four reaction mixtures, only one will add a dideoxy-terminator on to the primer sequence. If the mutation is present, it will be detected through the unique tag on the dideoxynucleotide after hybridization to the array. Multiple mutations can be simultaneously determined by tagging the DNA primer with a unique tag as well. Thus, the DNA fragments are reacted in four separate reactions each including a different tagged dideoxyterminator, wherein the tag is correlative with a particular dideoxynucleotide and detectable by non-fluorescent spectrometry, or potentiometry. The DNA fragments are hybridized to an array and non-hybridized material is washed away. The tags are cleaved from the hybridized fragments and detected by the

respective detection technology (*e.g.*, mass spectrometry, infrared spectrometry, potentiostatic amperometry or UV/visible spectrophotometry). The tags detected can be correlated to the particular DNA fragment under investigation as well as the identity of the mutant nucleotide.

The arrays of the present invention may be used to detect the products of single nucleotide extension assays (SNEAs). Two to 100 oligonucleotides may be immobilized per element in an array where each oligonucleotide in the element is a different or related sequence and complementary to a given product of single nucleotide extension assays (SNEAs). Preferably, each oligonucleotide sequence that is used to detect a SNEA is contained in adjacent elements. For example, the "A" product of a SNEA would be contained in element 1, the "T" product of a SNEA would be contained in element 2, the "C" product of a SNEA would be contained in element 3, the "G" product of a SNEA would be contained in element 4.

F. Oligonucleotide ligation assay

The oligonucleotide ligation assay (OLA). (Landegen et al., *Science 241*:487, 1988) is used for the identification of known sequences in very large and complex genomes. The principle of OLA is based on the ability of ligase to covalently join two diagnostic oligonucleotides as they hybridize adjacent to one another on a given DNA target. If the sequences at the probe junctions are not perfectly based-paired, the probes will not be joined by the ligase. The ability of a thermostable ligase to discriminate potential single base-pair differences when positioned at the 3' end of the "upstream" probe provides the opportunity for single base-pair resolution (Barony, *Proc. Natl. Acad. Sci. USA, 88*:189, 1991). When tags are used, they are attached to the probe, which is ligated to the amplified product. After completion of OLA, fragments are hybridized to an array of complementary sequences, the tags cleaved and detected by mass spectrometry.

Within one embodiment of the invention methods are provided for determining the identity of a nucleic acid molecule, or for detecting a selecting nucleic acid molecule, in, for example a biological sample, utilizing the technique of

oligonucleotide ligation assay. Briefly, such methods generally comprise the steps of performing amplification on the target DNA followed by hybridization with the 5' tagged reporter DNA probe and a 5' phosphorylated probe. The sample is incubated with T4 DNA ligase. The DNA strands with ligated probes are captured on the array by hybridization to an array, wherein non-ligated products do not hybridize. The tags are cleaved from the separated fragments, and then the tags are detected by the respective detection technology (*e.g.*, mass spectrometry, infrared spectrophotometry, potentiostatic amperometry or UV/visible spectrophotometry.

In the present invention, multiple samples and multiple mutations may be analyzed concurrently. Briefly, the method consists of amplifying the gene fragment containing the mutation of interest. The amplified product is then hybridized with a common and two allele-specific oligonucleotide probes (one containing the mutation while the other does not) such that the 3' ends of the allele-specific probes are immediately adjacent to the 5' end of the common probe. This sets up a competitive hybridization-ligation process between the two allelic probes and the common probe at each locus.

Within one embodiment of the invention methods are provided for determining the identity of a nucleic acid molecule, or for detecting a selecting nucleic acid molecule, in, for example a biological sample, utilizing the technique of oligonucleotide ligation assay for concurrent multiple sample detection. Briefly, such methods generally comprise the steps amplifying target DNA followed by hybridization with the common probe (untagged) and two allele-specific probes tagged according to the specifications of the invention. The sample is incubated with DNA ligase and fragments are captured on the arrays by hybridization. The tags are cleaved from the separated fragments, and then the tags are detected by the respective detection technology (*e.g.*, mass spectrometry, infrared spectrophotometry, potentiostatic amperometry or UV/visible spectrophotometry.

The oligonucleotide ligation assay as originally described by Landegren et al. (Landegen et al., *Science 241*:487, 1988) is a useful technique for the identification of sequences (known) in very large and complex genomes. The principle

of the OLA reaction is based on the ability of ligase to covalently join two diagnostic oligonucleotides as they hybridize adjacent to one another on a given DNA target. If the sequences at the probe junctions are not perfectly based-paired, the probes will not be joined by the ligase. The ability of a thermostable ligase to discriminate potential single base-pair differences when positioned at the 3' end of the "upstream" probe provides the opportunity for single base-pair resolution (Barony, *PNAS USA 88*:189, 1991). In the application of tags, the tags are attached to the probe, which is ligated to the amplified product. After completion of the OLA, fragments are hybridized to the array, the tags cleaved and detected by mass spectrometry.

In another embodiment, oligonucleotide-ligation assay employs two adjacent oligonucleotides: a "reporter" probe (tagged at the 5' end) and a 5'-phosphorylated/3' tagged "anchor" probe. The two oligonucleotides, which have incorporated different tags, are annealed to target DNA and, if there is perfect complementarity, the two probes are ligated by T4 DNA ligase. In one embodiment, the 3' tag is biotin and capture of the biotinylated anchor probe on immobilized streptavidin and analysis for the covalently linked reporter probe test for the presence or absence of the target sequences.

G. Other assays

The methods described herein may also be used to genotype or identification of viruses or microbes. For example, F+ RNA coliphages may be useful candidates as indicators for enteric virus contamination. Genotyping by nucleic acid amplification and hybridization methods are reliable, rapid, simple, and inexpensive alternatives to serotyping (Kafatos et. al., *Nucleic Acids Res. 7*:1541, 1979). Amplification techniques and nucleic aid hybridization techniques have been successfully used to classify a variety of microorganisms including *E. coli* (Feng, *Mol. Cell Probes 7*:151, 1993), rotavirus (Sethabutr et. al., *J. Med Virol. 37*:192, 1992), hepatitis C virus (Stuyver et. al., *J. Gen Virol. 74*:1093, 1993), and herpes simplex virus (Matsumoto et. al., *J. Virol. Methods 40*:119, 1992).

Genetic alterations have been described in a variety of experimental mammalian and human neoplasms and represent the morphological basis for the

sequence of morphological alterations observed in carcinogenesis (Vogelstein et al., *NEJM 319*:525, 1988). In recent years with the advent of molecular biology techniques, allelic losses on certain chromosomes or mutation of tumor suppresser genes as well as mutations in several oncogenes (*e.g.*, c-myc, c-jun, and the ras family) have been the most studied entities. Previous work (Finkelstein et al., *Arch Surg. 128*:526, 1993) has identified a correlation between specific types of point mutations in the K-ras oncogene and the stage at diagnosis in colorectal carcinoma. The results suggested that mutational analysis could provide important information of tumor aggressiveness, including the pattern and spread of metastasis. The prognostic value of TP53 and K-ras-2 mutational analysis in stage III carcinoma of the colon has more recently been demonstrated (Pricolo et al., *Am. J. Surg. 171*:41, 1996). It is therefore apparent that genotyping of tumors and pre-cancerous cells, and specific mutation detection will become increasingly important in the treatment of cancers in humans.

The following examples are offered by way of illustration, and not by way of limitation.

EXAMPLES

EXAMPLE 1

PREPARATION OF ARRAYING TIP FROM A COMMERCIAL SPRING PROBE.

This example describes the manufacture and modification of a spring probe tip for use in depositing samples in an array.

XP54P spring probes are purchased from Osby-Barton (a division of Everett Charles (Pomona, CA)). The probes are placed "tip-down" on an extra fine diamond sharpening stone and moved across the stone about 0.5 cm with gentle pressure. Approximately 0.005 inches (0.001 to 0.01 inches) of metal is removed from the end of the tip as observed by microscopy. The tip end is polished by rubbing the tip across a leather strip and then washed with water. Tips are stored dry or stored in 50% glycerol at -20°C. For use in preparation of arrays, the tips are mounted in a head in an

array fashion. The head is mounted on an robotic arm, which possesses controllable motion in the z-axis.

EXAMPLE 2

PREPARATION OF ARRAYS OF MICROSPHERES ON GLASS SLIDES.

Deposition of easily detectable microspheres on glass slides demonstrates reproducibility of array formation. In this procedure, a solution consisting of 56% glycerol. 0.01 M Tris pH 7.2, 5 mM EDTA, 0.01% sarkosyl, and 1% v/v Fluoresbrite Plain 0.5 μM microspheres (2.5% solids-latex), (Polysciences, Warrington, PA) is prepared. An arraying pin is submerged 5 mm into this solution for 5 sec. The microspheres are then repeatedly arrayed onto a glass slide. Photomicrographs of the slide are taken under fluorescence light using a filter for fluorescence. Figure 1 demonstrates that the amount of deposited solution in each area of the array is very consistent. Moreover, at least 100 deposits can be made per pickup that are virtually identical.

EXAMPLE 3

PREPARATION OF AN ARRAY USING A MODIFIED HYDROPHILIC SPRING PROBE

Sample pick-up, transfer and micro-droplet deposition is greatly enhanced when using a liquid transfer device that has a hydrophilic surface, especially when that device is a modified spring probe. Spring probes are rendered hydrophilic through the use of chemical agents acting to modify the surface of the probe or through coating the probe with a hydrophilic substance. In a preferred method, the tip of the spring probe is soaked in a 25 - 200 mM solution of 1,4 - dithiothreitol, 0.1 M sodium borate for 15 min to 2 hrs. Dithiothreitol reacts with gold surfaces through a thiol-gold coordination, which essentially hydroxylates the surface, making it hydrophilic.

An arraying solution is made consisting of 56% glycerol and 44% water colored with blue food color. The arraying tip is submerged 5 mm into the arraying

solution for 2 sec. The glycerol bearing tip is then robotically controlled to print 72 microspots in a 12x6 grid onto a silicon wafer. The spots produced were approximately 100-150 microns in diameter with 200 micron center to center spacing between spots. Figure 2 shows a CCD camera image of the grid produced. The standard deviation of spot diameter is approximately 15%.

EXAMPLE 4

COLORIMETRIC DETECTION OF ARRAYED OLIGONUCLEOTIDES.

Template oligonucleotide (75 µl of 0.5 µg/µl) (5'- hexylamine GTCATACTCCT-GCTTGCTGATCCACATCTG-'3) is reacted with 5 µl of a 20 mg/ml cyanuric chloride in 20 µl of 1 M sodium borate for 30 min at room temperature. From this reaction, an arraying solution is made, which consists of 56% glycerol, 56 ng/ul oligonucleotide, 0.06 mM sodium borate and 0.3 mg/ml cyanuric chloride. The arraying tip is submerged 5 mm into the arraying solution for 2 sec. The solution bearing tip is then robotically controlled to print 72 microspots in a 12x6 grid onto a polyethylenimine (PEI) coated silicon wafer. The spots produced are approximately 100-150 microns in diameter with 200 micron center to center spacing between spots. Following arraying, the unreacted PEI sites on the wafer are blocked with 100 mg/ml succinic anhydride in 100% n-methyl pyrrolinidone for 15 minutes followed by 3 washes in water. The unreacted cyanuric chloride sites are blocked with 0.1M glycine in 0.01 M Tris for 15 minutes with four washes in Tens buffer (0.1 M NaCl, 0.1% SDS, 0.01 M Tris, 5 mM EDTA). The template oligomer is then hybridized to its biotinylated complement (5'-Biotin-TGTGGATCAGCAAGCAGGAGTATG-3') for 20 min at 37°C followed by a wash in 6x Tens and 2x OHS (0.06 M Tris, 2 mM EDTA, 5x Denhardt's solution, 6x SSC [3 M NaCl, 0.3 M sodium citrate, pH 7.0], 3.68 mM N-lauroylsarcosine, 0.005% NP-40). The wafer is then soaked in 0.5 µg/ml alkaline phosphatase conjugated streptavidin for 15 min followed by a wash in 2x Tens, 4x TWS (0.1 M NaCl, 0.1% Tween 20, 0.05 M Tris). The microspots are then developed using Vector Blue (Vector Laboratories, Burlingame, California) (following kit protocol) and

imaged with a CCD camera and microscope. Figure 3 displays the image generated. The resulting microspots have approximately a 15% variation in diameter and intensity values varying approximately 10% as determined by NIH Image (National Institute of Health, Bethesda, MD).

EXAMPLE 5

MULTIPLE OLIGOS WITHIN A SINGLE ARRAY ELEMENT.

Two template oligos (#1, 5'-hexylamine-TGTGGATCAGCAAGCAGG AGTATG-3', #2 5'-hexylamine-ACTACTGATCAGGCGCGCCTTTTTTTTTTTTTTT TTTT-3') at 0.5 μg/μl are reacted separately with 5 μl of 20 mg/ml cyanuric chloride and 20 μl of 1M sodium borate in a total reaction volume of 100 μl for 30 minutes at room temperature. From these two reactions, arraying solutions are made of 56% glycerol and diluted combinations of the two reacted oligos (see Table below). Eight arraying tips are submerged 5 millimeters into each of the eight arraying solutions for 2 seconds. The solution bearing tips are robotically controlled to print two sets of eight 12x6 grids each containing 72 microspots onto a polyethylenimine (PEI) coated silicon wafer. Each grid represents a single arraying solution. The spots produced are approximately 100-150 microns in diameter with 200 micron center to center spacing between spots.

Following arraying, the unreacted PEI sites on the wafer are blocked with 100 mg/ml succinic anhydride in 100% n-methyl pyrrolinidone for 15 minutes with a 3x water wash. The unreacted cyanuric chloride sites are blocked with 0.1M glycine in 0.01 M Tris for 15 minutes with a 4x Tens (0.1 M NaCl, 0.1% SDS, 0.01 M Tris, 5 mM EDTA) wash. Two hybridizations are then carried out. In the first hybridization, one set of the eight arrayed oligo combinations is hybridized to the oligonucleotide, 5'-Biotin-TGTGGATCAGCAAGCAGGAGTATG-3', which is complementary to oligo #1. In the second hybridization, the other set of the eight arrayed oligo combinations is hybridized to the oligonucleotide (5'-BIOTIN-AAAAAA AAAAAAAAAAAAAAGGCGCGCCTGATCAGTAGT), which is complementary to

oligo #2. The hybridizations are conducted simultaneously under Hybriwell Sealing Covers (Research Products International Corporation, Mount Prospect, Illinois) for 20 minutes at 37°C followed by a 6x Tens, 2x OHS (0.06 M Tris, 2 mM EDTA, 5x Denhardt's solution, 6x SSC (3 M NaCl, 0.3 M sodium citrate, pH 7.0), 3.68 mM N-lauroylsarcosine, 0.005% NP-40) wash. Following hybridization, the wafer is soaked in 0.5 μg/ml horseradish peroxidase streptavidin for 15 minutes followed by a 2x Tens, 4x TWS (0.1 M NaCl, 0.1% Tween 20, 0.05 M Tris) wash. The microspots are then developed using 0.4mg/ml 4-methoxy 1-napthol (0.02%hydrogen peroxide, 12% methanol, PBS) with a final 3x water wash.

The set of mixed oligos that hybridize to the complement of oligo #1 show the greatest color intensity for the grid containing the highest concentration of oligo #1 and the least color intensity with the grid containing the lowest concentration of oligo #1. Whereas, the set of mixed oligos hybridized to the complement of oligo #2, showed the greatest color intensity for the grid containing the highest concentration of oligo #2 and the least color intensity with the grid containing the lowest concentration of oligo #2 (see figure 4).

Arraying Solution	Concentration of oligo #1 in arraying solution (ng/μl)	Concentration of oligo #2 in arraying solution (ng/μl)
1	56	0.44
2	28	0.88
3	14	1.8
4	7	3.5
5	3.5	7
6	1.8	14
7	0.88	28
8	0.44	56

From the foregoing it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.

CLAIMS

1. An array of nucleic acid molecules, comprising a solid substrate with a surface comprising discrete areas of nucleic acid molecules, wherein at least one area contains at least two nucleic acid molecules selected to have different sequence.

2. The array of claim 1, wherein there are less than 1000 discrete areas.

3. The array of claim 1, wherein the nucleic acids are oligonucleotides.

4. The array of claim 1, wherein each area contains at least two nucleic acid molecules with different sequence.

5. The array of claim 1, wherein at least one area contains from 2 to about 100 different nucleic acid sequences.

6. The array of claim 1, wherein an area is from about 20 to about 500 microns in diameter.

7. The array of claim 1, wherein a center to center distance between areas is from about 50 to 1500 microns.

8. The array of claim 1, wherein the nucleic acid molecules have known sequences.

9. The array of claim 1, wherein the nucleic acid molecules are covalently attached to the surface of the substrate.

10. The array of claim 9, wherein the covalent attachment is through an amine linkage.

11. The array of claim 10, wherein the surface of the substrate is coated with poly(ethyleneimine).

12. A method of hybridization analysis, comprising

(a) hybridizing labeled nucleic acid molecules to the array of nucleic acid molecules according to claim 1; and

(b) detecting label in areas of the array, therefrom determining which nucleic acid molecules on the array hybridized.

13. The method of claim 12, wherein at least one area of the array contains a known sequence and one of the labeled nucleic acid molecules is complementary to the known sequence.

14. The method of claim 13, wherein each area of the array contains the known sequence.

15. The method of claim 12, wherein the labeled nucleic acid molecules comprise nucleic acid molecules with different sequences, each carrying a different label.

16. The method of claim 12, wherein the label is selected from the group of radioactive molecules, fluorescent molecules, and cleavable mass-spec tags.

17. A method of identifying nucleic acid molecules in a sample, comprising:

(a) hybridizing labeled oligonucleotides to the nucleic acid molecules to form duplexes;

(b) isolating the duplexes;

(c) denaturing the duplexes;

(d) hybridizing the labeled oligonucleotides to the array of oligonucleotides according to claim 2, wherein the oligonucleotides on the array are complementary to the labeled oligonucleotides; and

(e) detecting label in areas of the array; therefrom identifying the nucleic acid molecules in the sample.

18. A method of identifying nucleic acid molecules in a sample, comprising:

(a) hybridizing oligonucleotides to the nucleic acid molecules;

(b) extending the oligonucleotide in the presence of a single labeled nucleotide to form duplexes;

(c) denaturing the duplexes;

(d) hybridizing the labeled oligonucleotides to the array of oligonucleotides according to claim 2, wherein the oligonucleotides on the array are complementary to the labeled oligonucleotides; and

(e) detecting label in areas of the array; therefrom identifying the nucleic acid molecules in the sample.

19. The method of claim 18, wherein the oligonucleotides in a discrete area of the array is complementary to an extension product of the nucleic acid molecules.

20. A method of identifying nucleic acid molecules in a sample, comprising:

(a) hybridizing at least two oligonucleotides to the nucleic acid molecules to form duplexes, wherein at least one oligonucleotide is labeled and the oligonucleotides hybridize to adjacent sequences on the nucleic acid molecules;

(b) ligating the oligonucleotides;

(c) denaturing the duplexes;

(d) hybridizing the labeled oligonucleotides to the array of oligonucleotides according to claim 2, wherein the oligonucleotides on the array are complementary to the ligated oligonucleotides; and wherein hybridization does not occur to unligated oligonucleotides; and

(e) detecting label in areas of the array; therefrom identifying the nucleic acid molecules in the sample.

21. A method of identifying mRNA molecules in a sample, comprising:

(a) hybridizing labeled oligonucleotides to the mRNA molecules to form duplexes;

(b) isolating the duplexes;

(c) denaturing the duplexes;

(d) hybridizing the labeled oligonucleotides to the array of oligonucleotides according to claim 2, wherein the oligonucleotides on the array are complementary to the labeled oligonucleotides; and

(e) detecting label in areas of the array; therefrom identifying the mRNA molecules in the sample.

22. The method of claim 21, wherein the mRNA molecules are isolated from cells treated with compounds suspected of being toxins.

23. The method of claim 22, wherein the oligonucleotides on the array are sequences of cytokines.

Visible light illumination:

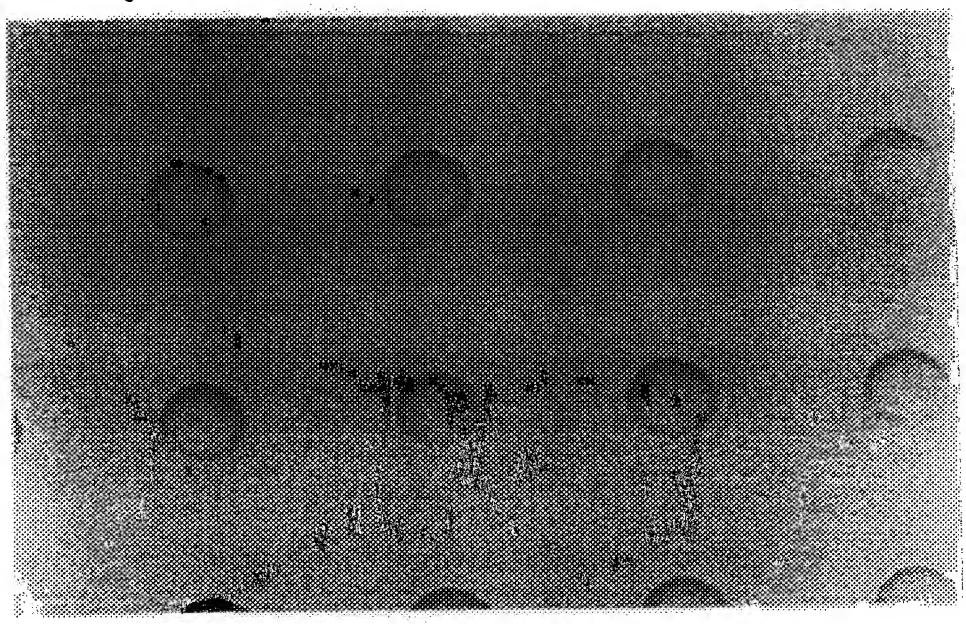

Fig. 1A

Fluorescence illumination:

Fig. 1B

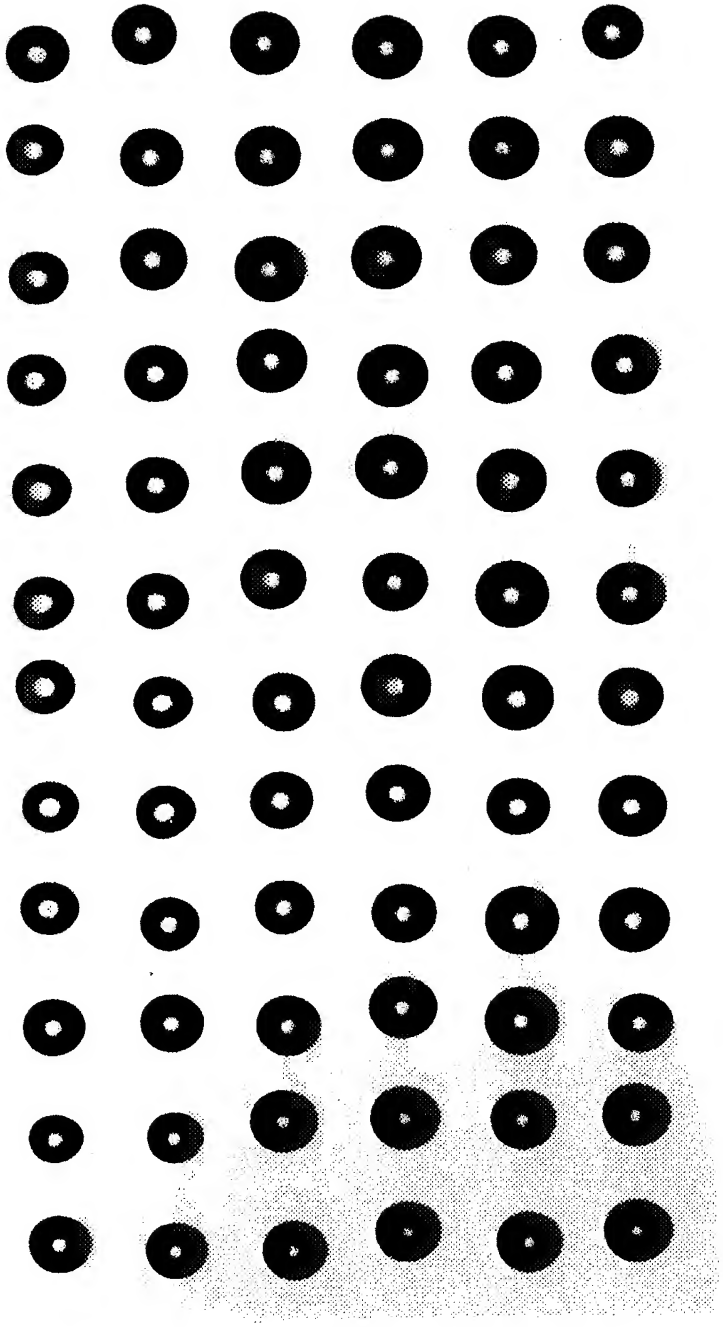

Fig. 2

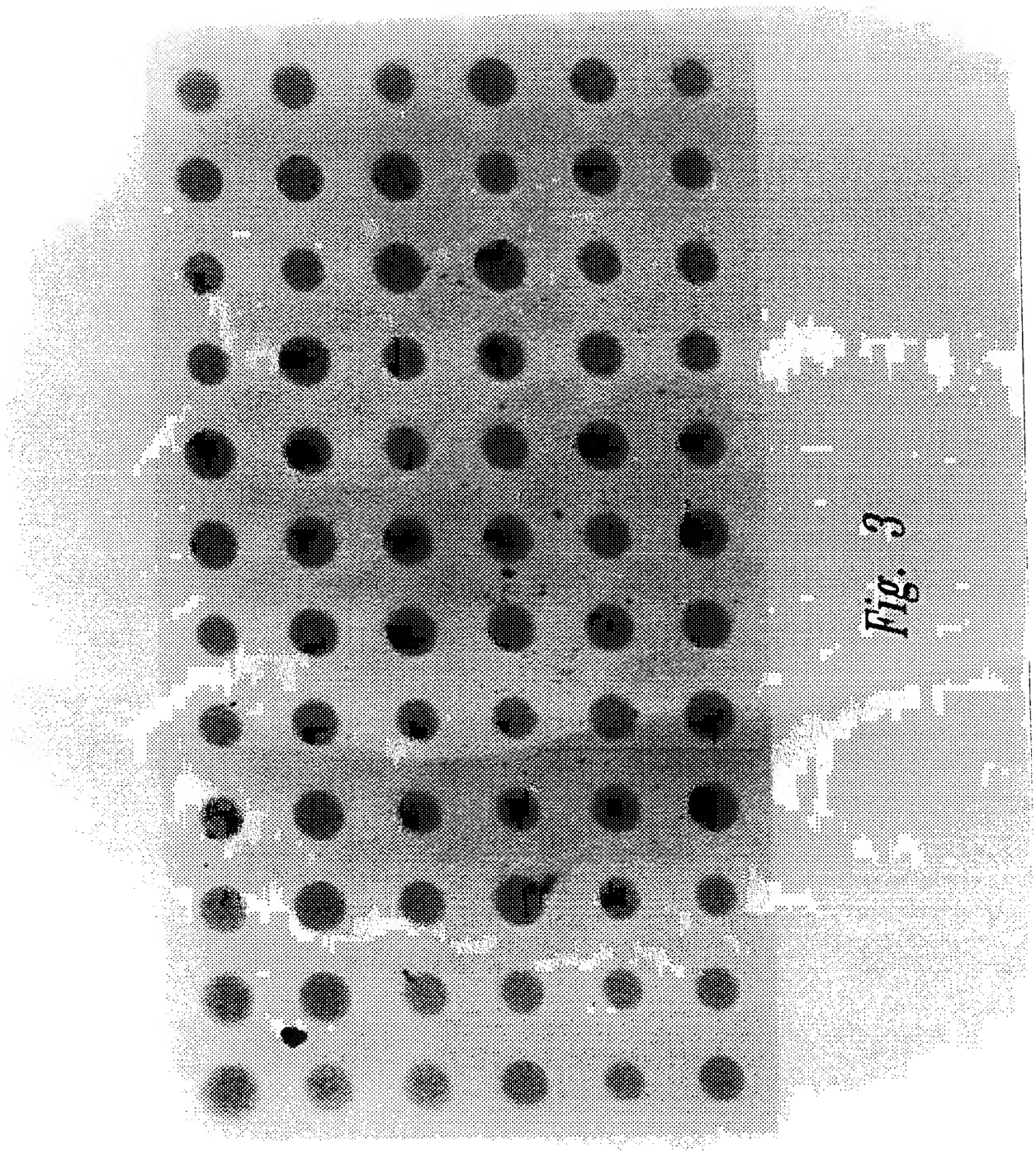

Fig. 3

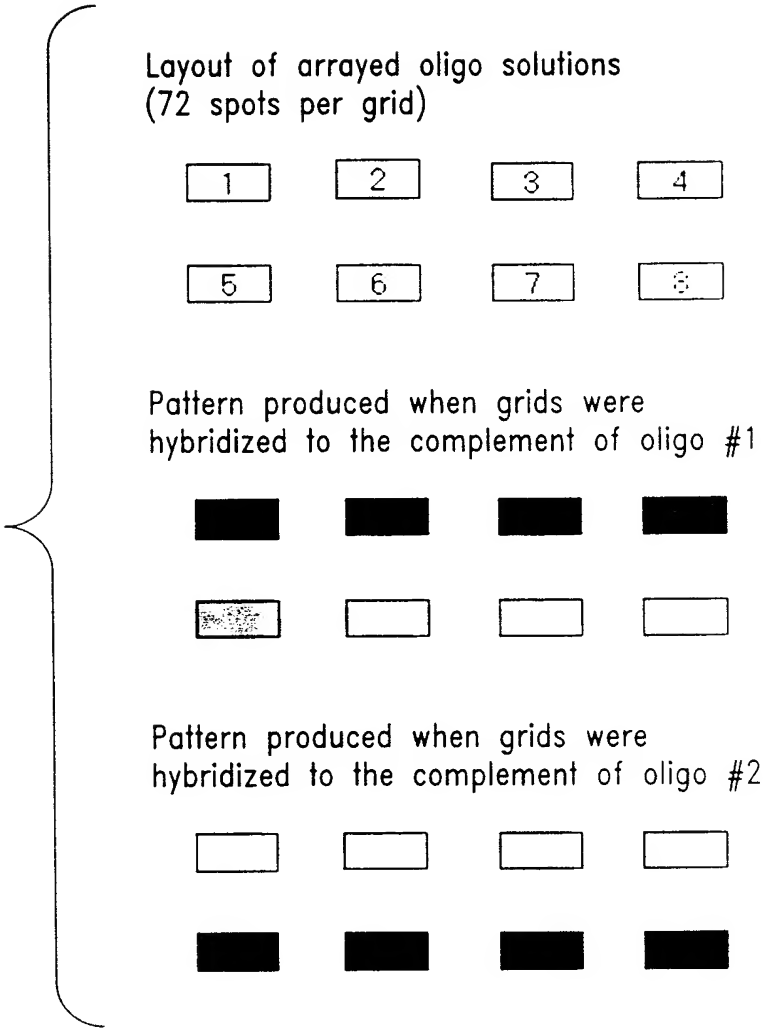
Layout of arrayed oligo solutions
(72 spots per grid)
1
2
3
4
5
6
7
8
Pattern produced when grids were
hybridized to the complement of oligo #1
Pattern produced when grids were
hybridized to the complement of oligo #2

Fig. 4

INTERNATIONAL SEARCH REPORT

Inte .ional Application No
PCT/US 98/15041

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	WO 96 27680 A (BOEHRINGER MANNHEIM GMBH ;KLEIBER JOERG (DE); OERUM HENRIK (DK); L) 12 September 1996 see whole document, esp. page 7, 2. para.; page 13, 1. para. and claims	1,3,5, 8-10,12, 13,15,16 18-22
Y	WO 95 20679 A (HYBRIDON INC) 3 August 1995 claims	20
Y	WO 90 13666 A (AMERSHAM INT PLC) 15 November 1990 claims	18,19
Y	WO 85 04720 A (FLOREY HOWARD INST) 24 October 1985 example 1 and claims	21,22

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
4 November 1998	11/11/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Müller, F

1

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

In ational Application No
PCT/US 98/15041

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document. with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	WO 96 31622 A (ISIS INNOVATION ;SOUTHERN EDWIN MELLOR (GB); PRITCHARD CLARE ELIZA) 10 October 1996 see the whole document	1-23
P,X	EP 0 785 280 A (AFFYMETRIX INC) 23 July 1997 see the whole document	1-5,8, 12,13,15

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int tional Application No
PCT/US 98/15041

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9627680 A	12-09-1996	DE 19548590 A	26-06-1997
		AU 5002996 A	23-09-1996
		CA 2214430 A	12-09-1996
		EP 0813610 A	29-12-1997
		WO 9627679 A	12-09-1996
		EP 0815259 A	07-01-1998
WO 9520679 A	03-08-1995	AU 1608095 A	15-08-1995
		CA 2180723 A	03-08-1995
		EP 0734457 A	02-10-1996
		JP 9508281 T	26-08-1997
		US 5637464 A	10-06-1997
WO 9013666 A	15-11-1990	CA 2045505 A	12-11-1990
		EP 0471732 A	26-02-1992
		JP 4505251 T	17-09-1992
WO 8504720 A	24-10-1985	AU 579631 B	01-12-1988
		AU 4231885 A	01-11-1985
		CA 1251119 A	14-03-1989
		DE 3583640 A	05-09-1991
		EP 0175776 A	02-04-1986
		JP 61501869 T	28-08-1986
WO 9631622 A	10-10-1996	EP 0820524 A	28-01-1998
EP 0785280 A	23-07-1997	NONE	